The terms "markup language" or "ML" refer to a language for special codes within a document that specify how parts of the document are to be interpreted by an application. In a word-processor file, the markup language specifies how the text is to be formatted or laid out, whereas in a particular customer schema, the ML tends to specify the text's structural function (e.g., heading, paragraph, etc.) The ML is typically supported by a word-processor and may adhere to the rules of other markup languages, such as XML, while creating further rules of its own.

The term "element" refers to the basic unit of an ML document. The element may contain attributes, other elements, text, and other building blocks for an ML document.

The term "tag" refers to a command inserted in a document that delineates elements within an ML document. Each element can have no more than two tags: the start tag and the end tag. It is possible to have an empty element (with no content) in which case one tag is allowed.

The content between the tags is considered the element's "children" (or descendants). Hence other elements embedded in the element's content are called "child elements" or "child nodes" or the element. Text embedded directly in the content of the element is considered the element's "child text nodes". Together, the child elements and the text within an element constitute that element's "content".

The term "attribute" refers to an additional property set to a particular value and associated with the element. Elements may have an arbitrary number of attribute settings associated with them, including none. Attributes are used to associate additional information with an element that will not contain additional elements, or be treated as a text node.

Illustrative Operating Environment

With reference to FIG. **1**, one exemplary system for implementing the invention includes a computing device, such as computing device **100**. In a very basic configuration, computing device **100** typically includes at least one processing unit **102** and system memory **104**. Depending on the exact configuration and type of computing device, system memory **104** may be volatile (such as RAM), non-volatile (such as ROM, flash memory, etc.) or some combination of the two. System memory **104** typically includes an operating system **105**, one or more applications **106**, and may include program data **107**. In one embodiment, application **106** may include a word-processor application **120** that further includes ML editor **122**. This basic configuration is illustrated in FIG. **1** by those components within dashed line **108**.

Computing device **100** may have additional features or functionality. For example, computing device **100** may also include additional data storage devices (removable and/or non-removable) such as, for example, magnetic disks, optical disks, or tape. Such additional storage is illustrated in FIG. **1** by removable storage **109** and non-removable storage **110**. Computer storage media may include volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information, such as computer readable instructions, data structures, program modules, or other data. System memory **104**, removable storage **109** and non-removable storage **110** are all examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by computing device **100**. Any such computer storage media may be part of device **100**. Computing device **100** may also have input device(s) **112** such as keyboard, mouse, pen, voice input device, touch input device, etc. Output device(s) **114** such as a display, speakers, printer, etc. may also be included. These devices are well know in the art and need not be discussed at length here.

Computing device **100** may also contain communication connections **116** that allow the device to communicate with other computing devices **118**, such as over a network. Communication connection **116** is one example of communication media. Communication media may typically be embodied by computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave or other transport mechanism, and includes any information delivery media. The term "modulated data signal" means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media includes wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media. The term computer readable media as used herein includes both storage media and communication media.

Word-Processor File Structure

FIG. **2** is a block diagram illustrating an exemplary environment for practicing the present invention. The exemplary environment shown in FIG. **2** is a word-processor environment **200** that includes word-processor **120**, ML file **210**, ML Schema **215**, and ML validation engine **225**.

In one embodiment, word-processor **120** has its own namespace or namespaces and a schema, or a set of schemas, that is defined for use with documents associated with word-processor **120**. The set of tags and attributes defined by the schema for word-processor **120** define the format of a document to such an extent that it is referred to as its own native ML.

Word-processor **120** internally validates ML file **210**. When validated, the ML elements are examined as to whether they conform to the ML schema **215**. As previously described above, a schema states what tags and attributes are used to describe content in an ML document, where each tag is allowed, and which tags can appear within other tags, ensuring that the documentation is structured the same way. Accordingly, ML **210** is valid when structured as set forth in arbitrary ML schema **215**.

ML validation engine **225** operates similarly to other available validation engines for ML documents. ML validation engine **225** evaluates ML that is in the format of the ML validation engine **225**. For example, XML elements are forwarded to an XML validation engine. In one embodiment, a greater number of validation engines may be associated with word-processor **120** for validating a greater number of ML formats.

FIG. **3** illustrates an exemplary ML file in accordance with aspects of the present invention. ML file **300** includes ML elements. An element in a markup language usually includes an opening tag (indicated by a "<" and ">"), some content, and a closing tag (indicated by a "</" and ">"). In this example, tags associated with ML include a "w:" within the tag (e.g., **302**). The "w:" prefix is used as shorthand notation for the namespace associated with the element.

The text contained within the document follows the "T" tag, making it relatively easy for an application to extract the text content from a word-processing document created in accordance with aspects of the invention. Given that the